THEORETICAL NOTES

The Robust Beauty of Ordinary Information

Konstantinos V. Katsikopoulos
Max Planck Institute for Human Development, Berlin, Germany, and Massachusetts Institute of Technology

Lael J. Schooler
Max Planck Institute for Human Development, Berlin, Germany

Ralph Hertwig
University of Basel

Heuristics embodying limited information search and noncompensatory processing of information can yield robust performance relative to computationally more complex models. One criticism raised against heuristics is the argument that complexity is hidden in the calculation of the cue order used to make predictions. We discuss ways to order cues that do not entail individual learning. Then we propose and test the thesis that when orders are learned individually, people’s necessarily limited knowledge will curtail computational complexity while also achieving robustness. Using computer simulations, we compare the performance of the take-the-best heuristic—with dichotomized or undichotomized cues—to benchmarks such as the naïve Bayes algorithm across 19 environments. Even with minute sizes of training sets, take-the-best using undichotomized cues excels. For 10 environments, we probe people’s intuitions about the direction of the correlation between cues and criterion. On the basis of these intuitions, in most of the environments take-the-best achieves the level of performance that would be expected from learning cue orders from 50% of the objects in the environments. Thus, ordinary information about cues—either gleaned from small training sets or intuited—can support robust performance without requiring Herculean computations.

Keywords: inductive inference, heuristics, take-the-best, robustness, bootstrapping

A psychological regularity touted as a potentially universal law of cognition is the effort–accuracy tradeoff: Only if people invest more cognitive effort do they stand to achieve more accuracy in their choices and judgments. More effort can take the form of searching for information exhaustively, spending plenty of time on the problem, or performing complex computations. Yet, this law may fall short of being universal. Inspired by the ideas of Herbert Simon, those in the research program on fast and frugal heuristics have developed computational models of heuristics as one precise interpretation of Simon’s (1956) notion of bounded rationality. Although computational models of heuristics were proposed before (e.g., Payne, Bettman, & Johnson, 1993; Tversky, 1972), they were built on the premise that people rely on heuristics because they lack the cognitive capacity to perform rational calculations or are willing to sacrifice accuracy by expending less effort. This view has been challenged by demonstrations that processes embodying bounded rationality, such as limited information search and noncompensatory processing, can lead to more accurate inferences than can be achieved by models based on more information and complex computations (Gigerenzer, Todd, & The ABC Research Group, 1999).

This challenge to the effort–accuracy tradeoff, however, has not remained undisputed. One important objection suggests that the heuristics’ success rides on complex computations. For example, many heuristics, such as take-the-best (Gigerenzer et al., 1999), do not use all available cues (i.e., variables that correlate, albeit imperfectly, with the criterion variable in a decision problem; see Brunswik, 1955) but instead order them, look them up one by one, and stop searching as soon as a discriminating cue is encountered. Juslin and Persson (2002, p. 575) stressed that much effort is needed to order cues. For example, in Gigerenzer et al.’s (1999) city-population-comparison task (e.g., which of the following two cities has more residents, Heidelberg or Bonn?), the computation of cue order involving 83 German cities and nine cues requires 30,627 comparisons (for details, see Juslin & Persson, 2002). Others have raised similar concerns (Dougherty, Franco-Watkins, & Thomas, 2008; Newell, 2005; Rakow, Hinvest, Jackson, & Palmer, 2004). Do simple heuristics really freeload on the effort hidden in the computation of cue orders? A first response to the “hidden effort” criticism is that a person can conveniently piggyback on the experiences of others, past or present. Evolution, culture, and other individuals can learn cue orders (for details, see Gigerenzer, Hofrage, &
Goldstein, 2008). Of course, despite evolution, culture, and the vicarious experience of others, sometimes a person finds herself with nobody to consult or imitate. This, however, does not dictate that the heuristics this person would use require massive effort for computing cue orders, for two reasons. First, some heuristics do without cue order. For example, the recognition heuristic (Goldstein & Gigerenzer, 2002) and the fluency heuristic (Hertwig, Herzog, Schoolder, & Reimer, 2008; Schoolder & Hertwig, 2005) need only one cue, benefiting from automatic, albeit potentially computationally complex, memory processes. Other heuristics, such as the priority heuristic, rely on risk preferences to order cues (Brandstätter, Gigerenzer, & Hertwig, 2006), and some heuristics bypass cues by relying on the frugal retrieval of exemplars (Juslin & Persson, 2002). A second reason why heuristics do not demand massive investment in the calculation of cue orders is that minute samples of knowledge can be used, which drastically restricts the computations required. Of course, to paraphrase Alexander Pope, a little cue knowledge can be a dangerous thing. It makes computations workable, but it may lead to ineffective cue orders. Whether severely limited knowledge is a curse or a blessing is the subject of our investigations.

We will test how well simple heuristics perform with ordinary cue information. In the context of this article, ordinary information refers to a person’s severely limited knowledge or a person’s knowledge of the sign of the correlation between a cue and a criterion, called the cue’s direction. Cue directions often follow people’s basic conceptions about how the world works: For example, cities with a high unemployment rate (cue) tend to have a high homelessness rate (criterion).

We present two investigations that combine computer simulations and human data. The first analyzes the performance of heuristics that use minute samples. The second analyzes the performance of heuristics that rely only on people’s intuitions about cue directions. We next describe the heuristics.

**Fast and Frugal Heuristics**

We focus on the inference models studied by Gigerenzer et al. (1999): three heuristics (tallying, take-the-best, and minimalist) and two benchmarks (linear regression and naïve Bayes). The task is a paired comparison, in which the goal is to predict which of two objects has the higher value on a criterion. For example, it can be asked which of two cities, Chicago or Los Angeles, has a higher homelessness rate. Lacking direct knowledge of these values, most people have to use cues. For example, the cues associated with a city’s homelessness rate could be the average temperature (a continuous cue) and the presence of a university (a binary cue).

These models differ in how they process cues. Linear regression weights and sums cues, predicting that the object with the higher sum has the higher criterion. Naïve Bayes selects the object with the higher probability of having the higher criterion value, given the objects’ values on the cues. In contrast, the three heuristics represent simplifications of weighted linear models. The first one, tallying (Dawes, 1979), dispenses with weighting and simply adds binary cue values. The take-the-best heuristic dispenses with adding and weighs cues simply by ordering them by a measure of cue goodness called validity and looking them up one at a time. As soon as a discriminating cue is found, search is stopped and a decision is made on the basis of that cue alone. Minimalist operates like take-the-best except that it randomly orders cues.

In sum, relative to the heuristics, the benchmarks use more complex computations and sophisticated information. Multiple regression and naïve Bayes use the precise values of regression weights and cue validities, whereas take-the-best uses the order of cue validities. Minimalist and tallying do not require cue validities or orders but only cue directions.

Czerlinski, Gigerenzer, and Goldstein (1999) applied these inference models to 20 data sets from domains such as biology, economics, and sociology. Each data set was split in half, and the parameters of each model (weights, validities, directions) were estimated on half of the data, the training set. The parameter estimates were used to predict, for each model, the paired comparisons on the other half of the data, the test set. This process was repeated 1,000 times. In the test set, the predictive accuracy of the models was as follows: minimalist 65%, regression with dichotomized cues 68%, tallying 69%, take-the-best with dichotomized cues 71%, naïve Bayes 73%, regression with undichotomized cues 76%, and take-the-best with undichotomized cues 76%.

Using a city-population-comparison task, Juslin and Persson (2002) and Chater, Oaksford, Nakisa, and Redington (2003) provided some evidence that heuristics have high predictive accuracy when the training set includes fewer than 50% of all objects. For example, Chater et al. tested take-the-best against a three-layer feed-forward connectionist network, two exemplar-based models, and a decision-tree-induction algorithm. With training sets of up to 40% of all objects, take-the-best outperformed or matched its competitors but was outperformed with larger training sets. Our first goal is to find out how robust these initial results are across many environments and with minute training set sizes with as few as two objects.

**Individual Learning From Minute Samples: How Good Could It Be?**

Table 1 lists the nine models included in the first simulation: two versions of multiple regression, one in which continuous cues were dichotomized (MRD) and one in which they were not (MRU); tallying (TAL) and minimalist (MIN), which both use dichotomized cues, and three versions of take-the-best, one that uses undichotomized cues (TTB) and two that use dichotomized cues. The latter two versions differ in their computation of validity, which can be done by a frequentist (TTBF) approach (in the undichotomized take-the-best version we used frequentist validity). We also included two versions of naïve Bayes: one with frequentist (NBF) and one with Bayesian (NBB) validity. The equation for the frequentist cue validity is \( v = r/(r + w) \), where \( r \) and \( w \) are the number of correct and incorrect comparisons, respectively, based on the cue. The frequentist approach may weigh the first observations too much, compromising performance on minute samples. The Bayesian approach overcomes this problem (Lee & Cummins, 2004) by calculating validity as \( v = (r + 1)/(r + w + 2) \).

---

1 This is a problem with some inference models, such as, for example, naïve Bayes. Naïve Bayes selects the object with the larger product of \( p(c_i|x) \) where \( v_i \) is the validity of any cue \( c_i \) that has a higher value on the object, and \( v_j \) is the validity of any cue \( c_j \) that has a higher value on the other object. If validities equal 0 or 1, the product will equal 0 for both objects (unless all cues with validity 1 point to one object and all cues with validity 0 point to the other), and naïve Bayes will have to guess. In line with this point, naïve Bayes with frequentist validity scored 50% for a training set of two objects (see Table 1).
Design

The predictive accuracy of the inference models was tested in 19 of the 20 data sets used by Czerlinski et al. (1999; we omitted a data set with fewer than 20 objects). To compare our results to previous work, we investigated a training set that included 50% of all objects. We also investigated training sets ranging in size from two to 10 objects. Averaged across the data sets, training sets of two and 10 objects equal 3% and 15% of all objects, respectively.

The models, with parameters (weights, validities, directions) estimated from all objects in the training set, were applied to the test set. When a model required binary cues, these were dichotomized cues using the sample median. The simulation was repeated 5,000 times for each model, environment, and training set size.

Results

Figure 1 plots mean predictive accuracy, defined as a model’s proportion of correct inferences in the test set as a function of the training set size. TTB_U performs best in all cases except when the training set has two objects. In this case, TTB_U is slightly behind TAL. For training sets with five to 10 objects, TTB_U surpasses the second-best model, NB_B, by at least 5%. With seven objects, TTB_U reaches 74%—near its maximum performance of 76%, achieved with a training set of 50% of all objects.

Bayesian validities guard against the extreme estimates that the frequentist approach may produce for very small training sets. Unlike take-the-best, naïve Bayes depends on point estimates of validity, so Bayesian validities could be important for its performance. With a training set of two objects, NB_B performs substantially better than does NB_P (58% vs. 50%; see also Footnote 1). This advantage disappears with 10 objects or more. The other benchmark, linear regression, does not fare well for very small training set sizes. This is not a new finding (Dawes, 1979; Hogarth & Karelaia, 2005). MRD does not outperform the MIN heuristic for any size of the training set. MR_U surpasses MIN only when there are six objects in the training set, and TAL only for nine objects in the training set.

In sum, we found that the models are differentially influenced by the information that minute samples afford. The predictive accuracy of the two benchmarks, linear regression and naïve Bayes, is compromised, presumably because training sets with 10 or fewer objects provide unreliable point estimates of regression weights and cue validities. Still, the use of Bayesian estimates provides a decent remedy for naïve Bayes. The heuristics, in contrast, seem capable of making do with very limited information: TAL is the most robust model for training sets with two objects, whereas TTB_U is the most robust model for sets including between three and 10 objects, with a difference in accuracy of more than 5 percentage points on average. In addition, with as few as seven objects TTB_U reaches a performance level close to the maximum achieved by any model with a 50% training set. Hogarth and Karelaia (2007) conjectured that heuristics such as TTB_U “would be less sensitive to sampling errors” (p. 751). This is exactly what we found. Why is it so?

The Robust Beauty of Cue Directions

One possible reason for the robustness of heuristics is that even minute samples identify a few good cues. For example, could it be that at least one of the, say, three most valid cues is ordered correctly? No. For training sets with two to eight objects, it is more likely that the three most valid cues are ranked incorrectly than that at least one of them is ranked correctly. Interestingly, getting the validity order right does not really matter much, as shown in Figure 2.

In the upper panel of Figure 2 we plot, for all training set sizes, the correlation between the predictive accuracy of take-the-best (and MIN) and how closely the validity order estimated from the training set corresponds to the validity order in the population. For each one of the 5,000 repetitions of the simulation and for each training set size, we measured the degree of correspondence between the two orders by a rank correlation. We separately considered the three most valid cues and all cues. None of the correlations between predictive accuracy and degree of correspondence reach a value higher than .2 for any training set size.

If cue order does not affect predictive accuracy much, then what does? Cue directions do. In the lower panel of Figure 2, we graph, for all training set sizes, the correlation between the predictive
accuracy of take-the-best (and MIN) and how closely the cue directions estimated from the training set correspond to the cue directions in the population. We measured the degree of correspondence between these two sets of cue directions by the number of cue directions in the population that are correctly estimated from the training set, for each one of the 5,000 repetitions of the simulation and for each training set size. As before, we considered the three most valid and all cues. The correlations start very high, between .6 and .7 for a training set with two objects. As the size of the training set increases, correlations decrease, possibly because there is increasingly less variability in the number of correctly estimated cue directions.

For all training set sizes, we calculated, using all data sets and all simulation repetitions per data set, the probability that the direction of a cue was estimated correctly. To avoid overestimating this probability, we used all repetitions, not only those in which a cue discriminated and its direction could be judged. Even so, the probability is high. By sampling two objects, cue direction is estimated correctly with a probability of about .6, surpasses .7 for a training set with four objects, .51 for a set with eight objects, and .56 for a set with 10 objects.

In sum, minute samples do not veridically represent the population validity order. Yet, they afford an ordinary but informative piece of information—the direction of a cue. Moreover, even minute samples offer a slightly better-than-chance probability of looking up good cues first. This, together with good cue directions, explains the impressive robustness of take-the-best.

**Cue Directions: Individual and Collective Intuitions**

Our first investigation established that fast and frugal heuristics do not necessarily need good cue orders, relying instead on good cue directions to make accurate inferences. In our second investigation, we ask: How good or bad are people’s intuitions about cue directions? The father of the TAL heuristic, Robyn Dawes (1979), wrote that “people are good at picking out the right predictor variables and at coding them in such a way that they have a conditionally monotone relationship with the criterion” (p. 573). But it is known that individuals often develop false beliefs about how variables are associated. For example, Chapman and Chapman (1969) found that students perceived positive correlations between people’s interpretations of ink blots and homosexuality, although in reality the correlation in the experimental material was negative.
We empirically probed people’s intuitions about the directions of cues on 10 of the data sets used before. To see whether these intuitions can lead heuristics to perform well, we also tested the performance of bootstraps of heuristics. That is, we used the cue directions judged by participants as input to take-the-best and TAL. Bootstrapping may refer to using the judgments of a person as input to the model of the same person. We use the term more broadly, including the use of an aggregate judgment of a group (defined later) as input.  

Method

Of the data sets used previously, we eliminated some that are obscure (e.g., predicting rainfall from cloud seeding) or strongly overlap with others (e.g., we selected predicting obesity instead of percentage of body fat). Using the same constraints, we reduced the cues to an average of 5.1 cues per data set, varying from three to eight, ending up with a total of 51 continuous and binary cues. Table 2 lists the remaining environments and cues. For two environments, participants were likely to have some direct experience (German city populations and body fat in men). It is unlikely, however, that our German participants had experience with the other environments, such as the housing market in Erie, Pennsylvania.

Rather than asking our participants, 50 students from three universities in Berlin, to state the sign of a correlation between cue and criterion, we asked them whether an increase in the value of a specific cue would be associated with an increase or a decrease in

---

2 We are not using the term bootstrap to refer to the statistical procedure based on resampling.
We also asked them to rate the criterion value. We also asked them to order cues by how useful they thought the cues would be. The order in which environments were presented across participants. For some cues, some participants did not answer, but we collected more than 45 judgments for each cue.

### How Good Are People’s Intuitions About Cue Directions?

Across all 10 environments, people judged cue directions correctly in 67% of all cases. The directions of about half of the cues (27 out of 51) were judged correctly by more than 85% of participants. Directions of about three fourths of the cues (38 out of 51) were judged correctly by more than 50% of the participants. Cues judged incorrectly by more than 50% of the participants occur mostly in three environments: mammals' sleep (four cues), highway accidents (three cues), and fuel consumption (three cues). People’s incorrect intuitions are not surprising. For example, the number of drivers with a license in a state is negatively correlated with the average fuel consumption in the state. Other cue directions may be justifiable in retrospect but are still surprising, such as, for example, the number of car accidents on a road segment being negatively correlated with the speed limit. Finally, some cue directions simply require expert knowledge. For instance, it is not clear, to dilettantes in biology such as the authors, why the gestation index is negatively correlated with how much a mammal sleeps.

### How Well Do Heuristics Perform When Fed With People’s Cue Directions?

To answer this question, we constructed and tested bootstrap models of take-the-best and a benchmark that we call calibrated take-the-best model. The predictive accuracy of the calibrated take-the-best is the predictive accuracy of TTB as analyzed in the first simulation. Across the 10 environments and 51 cues listed in Table 2, TTB was calibrated on a training set with 50% of all objects in each environment. The predictive accuracy of the first bootstrap, which we call individual take-the-best, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions and cue orders probed from one of the 50 participants. The second bootstrap, social take-the-best, used the cue directions of the majority of participants (García-Retamero et al., in press). To determine cue order, social take-the-best exploits the strength of the majority about cue directions. For example, if 60% of participants agreed on the direction of one cue, whereas 70% of participants agreed on the direction of another cue, the second cue would be ranked higher.

We also conducted the corresponding analysis for TAL, testing two bootstrap models and a benchmark, calibrated TAL. The predictive accuracy of the first bootstrap, individual TAL, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions of one of the 50 participants. The second, social TAL, takes the direction of a cue to be the direction judged by the majority of the participants (García-Retamero et al., in press).

Table 3 shows the performance of the three take-the-best and three TAL models. Across all 10 environments, the bootstrap models lag from 7 to 10 percentage points behind the calibrated model. Table 2 also shows how accuracy depends on the environment. We distinguish post hoc between counterintuitive environments, in which cue directions proved difficult for people to judge (the mammals’ sleep, highway accidents, and fuel consumption environments), and intuitive environments, in which cue directions were easier to judge (the other seven environments). In counterintuitive environments, for both take-the-best and TAL, the calibrated model does much better than the bootstrap models. In intuitive environments, however, the bootstrap models, especially the social bootstraps, do well and catch up or even outperform the calibrated models of take-the-best and a benchmark that we call calibrated take-the-best model. The predictive accuracy of the calibrated take-the-best is the predictive accuracy of TTB as analyzed in the first simulation. Across the 10 environments and 51 cues listed in Table 2, TTB was calibrated on a training set with 50% of all objects in each environment. The predictive accuracy of the first bootstrap, which we call individual take-the-best, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions and cue orders probed from one of the 50 participants. The second bootstrap, social take-the-best, used the cue directions of the majority of participants (García-Retamero et al., in press). To determine cue order, social take-the-best exploits the strength of the majority about cue directions. For example, if 60% of participants agreed on the direction of one cue, whereas 70% of participants agreed on the direction of another cue, the second cue would be ranked higher.

We also conducted the corresponding analysis for TAL, testing two bootstrap models and a benchmark, calibrated TAL. The predictive accuracy of the first bootstrap, individual TAL, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions of one of the 50 participants. The second, social TAL, takes the direction of a cue to be the direction judged by the majority of the participants (García-Retamero et al., in press).

Table 3 shows the performance of the three take-the-best and three TAL models. Across all 10 environments, the bootstrap models lag from 7 to 10 percentage points behind the calibrated model. Table 2 also shows how accuracy depends on the environment. We distinguish post hoc between counterintuitive environments, in which cue directions proved difficult for people to judge (the mammals’ sleep, highway accidents, and fuel consumption environments), and intuitive environments, in which cue directions were easier to judge (the other seven environments). In counterintuitive environments, for both take-the-best and TAL, the calibrated model does much better than the bootstrap models. In intuitive environments, however, the bootstrap models, especially the social bootstraps, do well and catch up or even outperform the calibrated models of take-the-best and a benchmark that we call calibrated take-the-best model. The predictive accuracy of the calibrated take-the-best is the predictive accuracy of TTB as analyzed in the first simulation. Across the 10 environments and 51 cues listed in Table 2, TTB was calibrated on a training set with 50% of all objects in each environment. The predictive accuracy of the first bootstrap, which we call individual take-the-best, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions and cue orders probed from one of the 50 participants. The second bootstrap, social take-the-best, used the cue directions of the majority of participants (García-Retamero et al., in press). To determine cue order, social take-the-best exploits the strength of the majority about cue directions. For example, if 60% of participants agreed on the direction of one cue, whereas 70% of participants agreed on the direction of another cue, the second cue would be ranked higher.

We also conducted the corresponding analysis for TAL, testing two bootstrap models and a benchmark, calibrated TAL. The predictive accuracy of the first bootstrap, individual TAL, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions of one of the 50 participants. The second, social TAL, takes the direction of a cue to be the direction judged by the majority of the participants (García-Retamero et al., in press).

Table 3 shows the performance of the three take-the-best and three TAL models. Across all 10 environments, the bootstrap models lag from 7 to 10 percentage points behind the calibrated model. Table 2 also shows how accuracy depends on the environment. We distinguish post hoc between counterintuitive environments, in which cue directions proved difficult for people to judge (the mammals’ sleep, highway accidents, and fuel consumption environments), and intuitive environments, in which cue directions were easier to judge (the other seven environments). In counterintuitive environments, for both take-the-best and TAL, the calibrated model does much better than the bootstrap models. In intuitive environments, however, the bootstrap models, especially the social bootstraps, do well and catch up or even outperform the calibrated models of take-the-best and a benchmark that we call calibrated take-the-best model. The predictive accuracy of the calibrated take-the-best is the predictive accuracy of TTB as analyzed in the first simulation. Across the 10 environments and 51 cues listed in Table 2, TTB was calibrated on a training set with 50% of all objects in each environment. The predictive accuracy of the first bootstrap, which we call individual take-the-best, is the average of the predictive accuracies of 50 bootstraps, each using the cue directions and cue orders probed from one of the 50 participants. The second bootstrap, social take-the-best, used the cue directions of the majority of participants (García-Retamero et al., in press). To determine cue order, social take-the-best exploits the strength of the majority about cue directions. For example, if 60% of participants agreed on the direction of one cue, whereas 70% of participants agreed on the direction of another cue, the second cue would be ranked higher.

3 We also implemented a version of social take-the-best that uses people’s intuitions about not only cue directions but also cue orders. It uses Borda counts, the sum of the ranks that a cue was given by all participants, as a proxy for cue ranks (García-Retamero et al., in press). This model reached essentially the same predictive accuracy as did social take-the-best with cue directions and the majority rule.
General Discussion
In what follows, we first discuss why getting cue directions right appears to be so crucial for the performance of fast and frugal heuristics, and then we turn to the role of causal theories underlying intuitions about cue directions.

The Robust Beauty of Ordinary Information
The title of our article alludes to research by Dawes, Hastie, and Kameda on the “robust beauty” of TAL and its group analogue, the majority rule (Dawes, 1979; Hastie & Kameda, 2005). Consistent with the results of their work, the success of TAL in our simulations shows that aggregating cue directions is a robust strategy. What is puzzling about the remarkable performance of take-the-best with undichotomized cues is that it dispenses with aggregation. It relies on just the direction of literally a single cue, the one with the highest validity estimate on the basis of the training set. In the case of continuous cues, this cue will almost always discriminate between two objects. A full solution to this puzzle is still lacking, but we can gain some insight by using two approaches taken in the past to explain the strong performance of heuristics.

First, heuristics achieve high accuracy if the cue structure is noncompensatory. There are at least two meanings of a noncompensatory cue structure, which have been studied analytically: Katsikopoulos and Martignon (2006) showed that, essentially, if the cue with the highest validity has a much higher validity than do all other cues, then take-the-best with binary cues has maximum accuracy among all possible inference models. And Hogarth and Karelaia (2005) showed that, essentially, if the variance in criterion values accounted for by the cue that take-the-best looks up first is higher than the variance in criterion values accounted for by using all cues in a linear regression (adjusting for the total number of cues), then take-the-best with undichotomized cues is more accurate than regression is.

We offer a speculation for the performance of take-the-best with undichotomized cues rooted in how dichotomization affects validity. One might expect the validity of a continuous cue, under conditions to be specified, to decrease when the cue is dichotomized. The values of two objects on a continuous cue can (a) both lie on the same side of the threshold used to dichotomize the cue or (b) lie on the opposite sides of this threshold. In (b), the validity of the cue is unaffected by dichotomization. In (a), a dichotomized cue does not discriminate, whereas an undichotomized cue does. For environments where the validity of a continuous cue in (a) is higher than its validity in (b), dichotomization lowers cue validity. Irwin and McClelland (2003) made similar observations about the statistical power lost from dichotomizing continuous cues.

In a second approach to explaining the strong performance of take-the-best, Gigerenzer and Brighton (2009) used the fact that the predictive accuracy of any inference model is the sum of its predictions’ bias and variance. They conjectured that heuristics achieve high predictive accuracy because their comparatively low variance compensates for their comparatively high bias. Extrapolating from this account, we suspect that take-the-best with dichotomized cues will have lower variance in its predictions than will take-the-best with undichotomized cues. When cues are continuous, take-the-best determines the relative ranks of two objects, according to the criterion, on the basis of the first cue only. When cues are dichotomized, however, the ranking of objects can, in the most extreme case, depend on the complete order of cues. That is, with three continuous cues there would be, in principle, three rankings of the objects, whereas with three dichotomized cues there would be, in principle, six (3!). This difference could decrease the variance in take-the-best’s predictions with continuous cues, relative to dichotomous cues.

Cue Directions: Occam’s Razor and Causal Theories
Entities must not be multiplied beyond necessity. If so, what is gained by explaining a person’s judgment for one kind of binary choice, say, which of two mountains in the Swiss Alps is higher, on the basis of another binary variable, intuitions about cue direction? Clearly the most direct solution to the inference problem is to retrieve the mountains’ height directly from memory (see the concept of “local mental models” in Gigerenzer, Hoffrage, & Kleinböltung, 1991). However, there will be situations in which there is no direct knowledge about the criterion. Then, a person may vicariously substitute direct knowledge with probabilistic cues. Our analyses have shed light on the key role of cue directions in this process of substitution.

Table 3
Predictive Accuracy (%) of the Models in the Simulation in the Second Investigation

<table>
<thead>
<tr>
<th>Model</th>
<th>All 10 environments</th>
<th>3 Counterintuitive environments</th>
<th>7 Intuitive environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take-the-best</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Individual</td>
<td>58</td>
<td>42</td>
<td>65</td>
</tr>
<tr>
<td>Social</td>
<td>60</td>
<td>42</td>
<td>68</td>
</tr>
<tr>
<td>Tallying</td>
<td>67</td>
<td>69</td>
<td>66</td>
</tr>
<tr>
<td>Individual</td>
<td>59</td>
<td>46</td>
<td>64</td>
</tr>
<tr>
<td>Social</td>
<td>60</td>
<td>41</td>
<td>68</td>
</tr>
</tbody>
</table>

Note: The predictive accuracy of a calibrated model and two bootstrap (called individual and social) models, separately for take-the-best and tallying. The calibrated models use all cue-related information in a training set with 50% of all objects in the population, whereas the bootstrap models use cue direction information provided by 50 people (see text for details). Results are provided across all environments and separately for those environments where cue directions were found to be difficult for people to judge (counterintuitive environments) and easy for people to judge (intuitive environments).
We speculate that people could arrive at intuitions about cue directions on the basis of their causal knowledge about how the world works. According to Waldmann, Hagmayer, and Blaisdell (2006), people have a “natural capacity to form causal representations” (p. 310), based, in part, on regularities such as “causes typically temporally precede their effects” (p. 308). García-Retamero, Hoffrage, and Dieckmann (2007) have argued that causal representations help people focus on causal associations between cues and criteria, which are likely to be robust. They demonstrated that people were faster to learn causal associations, relative to equally valid associations that were not presented to participants as causal, and that they were more likely to search for cues that were causally related to criteria.

Tentative evidence that causal theories shaped our participants’ intuitions about cue directions can be found in one of the common errors in judged cue directions. For inferring the rate of highway accidents, people could causally link speed limit (cue) to lower accident rates (criterion): “Low speed increases the possibility to respond in time when necessary, and thus the possibility to avoid a collision becomes higher.” Counterintuitively, however, lower speed limits in the present data set are associated with more accidents. That is, causal theories can be and sometimes are wrong—but they nevertheless give rise to intuitions about cue directions.

**Conclusion**

The effort-accuracy tradeoff carries the ring of a general law of cognition: Investing less effort is tantamount to achieving lower accuracy. Challenging the belief that accuracy and effort inescapably trade off, research on fast and frugal heuristics has demonstrated that less information and computation can yield better performance. Countering the argument that heuristics’ success rides on the effort put into calculating cue validities and orders, we showed that information limitations that reduce effort do not always hurt accuracy. Simple heuristics can be robust even if simplicity is secured through ordinary information about cue directions, garnered from limited knowledge or found in people’s intuitions.

**References**


Received July 1, 2009
Revision received March 18, 2010
Accepted March 19, 2010